

## Executive Summary

The Office of Cooperative Threat Reduction (CTR) in the U.S. State Department's Bureau of International Security and Nonproliferation (ISN) conducts programming to prevent both proliferator states and non-state terrorist groups from developing or acquiring weapons of mass destruction (WMD) and their delivery systems. CTR measures the progression of security related to high-risk biological, chemical, and nuclear material—as well as the security culture—dedicated to countering proliferation in CTR partner countries. CTR's WMD programmatic efforts include the Biosecurity Engagement Program (BEP), the Chemical Security Program (CSP), the Partnership for Nuclear Threat Reduction (PNTR), as well as the Iraq Team. Additionally, CTR works to constrain states and state-sponsored actors seeking to engage in proliferation efforts and apply pressure on them to deter such activities associated with North Korea, Iran, China, Russia, and Syria. This work is referred to in this report as counterproliferation (CP) engagement.

This evaluation report provides CTR with a comprehensive analysis to assess program effectiveness and inform future activities. The evaluation team implemented the existing CTR evaluation tool with the following objectives:

1. Offer options to augment CTR's operations with methodologies to improve outcome measurement in complex operating environments.
2. Enable CTR to obtain a refined understanding of the effectiveness of its programming and related challenges.
3. Report lessons and best practices from data collected and provide CTR with case studies and examples to show impact.

## Methodology

The evaluation tool used by CTR applies a mixed-methods approach to capture both qualitative and quantitative trends in the evolution of security within a specified set of countries at different levels of analysis: institutional, capacity, and national. National metrics capture the overall legal and regulatory framework that provides the structural foundation for security around nuclear, biological, and chemical materials. The institutional level measures the degree to which government institutions have adopted security measures and promoted a security culture. The capacity metric measures broader security components within countries where CTR engages, such as law-enforcement activities, professional societies, and training. Lastly, the counterproliferation metric examines regulations directed at reducing the illicit procurement of resources (financial, intellectual, and material) by proliferating states and non-state groups.

The national and counterproliferation metrics are based on open-source research that informed survey questions and scores on a one-to-five scale. The capacity and institutional metrics use the same scoring framework, based on surveys that were sent to implementers and program officers. This evaluation tool also integrates a review of After-Action Reports (AARs) from country-specific CTR-funded programs. This report also provides a set of frameworks to serve as a basis for the development of new indicators and metrics for CTR's new activities focusing on Russia and China. The evaluation team reviewed CTR's logic model and numerous

Notices of Funding Opportunity (NOFOs), and conducted online research to inform these frameworks and provide illustrative metrics in each area.

## Key Observations

The final section of the report provides details on key, country-level, and programmatic observations. While it makes sense to separate these two types of observations for presentation purposes, our key observation is that programmatic capacity to collect data and manage knowledge systemically determines the degree and quality of a program's ability to collect data on country-level trends and understand the program's contributing role.

At the country level, analysis shows continued areas of focus, such as strengthening coordination and communication systems. Implementer feedback indicates a fragmentation of knowledge-sharing across agencies as well as a lack of clarity on roles and responsibilities and Standard Operating Procedures (SOPs). Security awareness is another ongoing issue, and raises a need for capacity building in labs, facilities, professional societies, and government agencies.

At the programmatic level, the team found that the evaluation tool provides a set of standardized parameters that allow easy, presentable, and digestible data. However, the knowledge gathered from the survey (and transcribed into scoring) has two key limitations:

1. It does not help elucidate what the numbers or decimal changes really mean or the degree to which CTR's engagements contributed to changes; and
2. The standardization of scoring across the multitude of activities, implementers, and surveys broadens the scope, but diminishes and diffuses the quality and focus of learning coming from the field as data collection (and therefore, results) do not match priorities driven by a learning agenda. Furthermore, the program's capacity to monitor, evaluate, and learn about its progress- let alone its impact (especially for an office as rich and diverse as CTR)-is inherently limited by its sole reliance on an annual blanket survey.

## Summary of Findings

### *National*

At the national level, scores increased slightly over last year across the board in nearly every discipline (biological, chemical, and nuclear) in all countries with CTR programming. In 2020, the overall scores of a third of the countries (six) in the portfolio were unchanged compared to 2019. Since last year, 12 countries in the portfolio saw an increase in their overall scores, averaging 2.43 on a one-to-five point scale, reflecting an increase of 0.02. Bangladesh, Egypt, Indonesia, Pakistan, and Iraq have registered the biggest changes across disciplines. Overall, the analysis shows movement towards regulatory transparency, continual improvement in security culture, and increased training offerings in countries. The section of this report on national metrics also includes country spotlights on Egypt, Pakistan, Bangladesh, and Libya, to put their score ranges in the context of the activities reviewed.

## Capacity

The highest-scoring countries in terms of capacity were Kenya, Malaysia, Egypt, and India. As discussed in the capacity metrics section of this report, fewer engagements-particularly those related to law enforcement and professional societies-occurred in the field in 2020 due to the COVID-19 pandemic, resulting in a smaller sample size than in previous years. The pre-/post- section of the training survey, a component of the capacity metrics, showed an average 26 percent improvement among participants across the three disciplines. Analysis also shows that the ability to capture long-term change after workshops and training was limited, as many respondents marked "N / A" or "I don't know" in response to questions about changes in areas such as advances in inventory management, internal policy adjustment, or operational security procedures. This could be the result of lower levels of engagement with participants.

## Institutional

The small sample size posed a challenge in the institutional category, as input was received from fewer facilities as a result of fewer site visits (five this year compared to 46 last year) due to COVID-19 restrictions. Based on a review of five facilities in the biological discipline, Turkey scored highest (4.51), Algeria scored just above average at 3.12, and the Democratic Republic of the Congo (DRC) scored lowest (1.31). The highest scores were in Managerial Responsibility for Promotion of Security Culture and Material Protection, Control, and Accountability (MPC&A), and the lowest was in Personnel Reliability. In the chemical discipline, Turkey scored 4.68 and Lebanon scored 3.23. The highest scores were in MPC&A, Security Culture, and Managerial Responsibility at 4.28, 3.43, and 3.4, respectively.

## Counterproliferation

The counterproliferation metric used the same methodology as in 2019. It sought to rate countries in three standard categories: policy issues, threat/risk/vulnerability assessment, and regulatory authority. Overall, between 2019 and 2020, CP scores increased by 12 percent, from an average raw score of 2.98 to 3.31. Of the three standard categories, the largest relative increase (19 percent, from 1.86 to 2.22) was in policy issues, followed by threat/risk/vulnerability assessment (10 percent, from 3.99 to 4.37), and regulatory authority (eight percent, from 3.10 to 3.35).

## Limitations

**Open-Source/Availability of Information:** As scoring relies on open sources (for national and counterproliferation metrics), access to relevant documentation remains a challenge and varies based on the level of transparency of government ministries as well as other entities and databases.

**Scoring Methodology:** The evaluation team used a five-point scoring scale and sheet consisting of definitions and descriptors that were developed in the previous evaluation period. The sheet was used in all the countries to answer survey questions related to both national metrics and counterproliferation. However, the scoring necessarily reflects some level of reviewers' subjective judgement, as open-source documentation found in the review did not always directly answer the questions for each discipline. Second, other metrics were converted to a five-point scale to make scoring consistent across the evaluation. However, as there is no guidance for the meaning of a score in the context of the capacity metrics and their components,

interpretational inconsistencies may account for some of the variation in scores compared to last year. Finally, those inconsistencies made it challenging to provide a cross-year comparison and establish the overall capacity metrics per country.

**Sample Criteria:** Survey links for each type of engagement (corresponding to the various metrics) were sent to a sample of 330 program implementers and program officers who engaged directly with foreign partners and were identified by CTR as survey respondents.

Out of the 330 surveys sent out, QED received 241 responses. Of these respondents, 68 individuals opted in to take the survey for CP, six took the institutional survey, seven took law enforcement, eight took professional societies, and 145 underwent training; altogether, this disproportionate mix of engagement types reduces the comparability of the findings.

**Bias:** The different types of surveys were presented as a package for responders to choose the type of engagements on which to report. However, the survey results could be affected by selection bias, as some unobservable, inherent quality about the individuals that led them to opt into specific surveys may be present, ultimately providing answers not necessarily representative of the larger population.

**Alignment of Engagement Scope and Survey Components:** Many survey responses were marked "N / A" or "I don't know," indicating a possible misalignment between the scope of engagements and the survey questions. For example, analysis of the training survey showed that the sections of the surveys that were completely filled out were the ones pertaining strictly to the immediate pre-/post- sections, while the parts of the survey related to longer-term changes and aspects of security associated with institutional changes were only partially completed. Similarly, given the nature of site-visit engagement, several of the categories in the survey related to institutional metrics may be outside their purview, as noted in previous years' reports. Site visits ranged in scope and type of observation.

## Guidance for Future Actions

With its vast network of implementers, wide range of engagements, and diverse institutional memory, CTR is well-positioned to orient its monitoring and evaluation (M&E) activities around structured and purposeful programmatic learning. A learning strategy can guide implementers to provide knowledge to effectively inform CTR's learning needs and decision-making. This approach also lends itself to the creation of internally and externally focused learning activities, such as portfolio reviews, that allow for continuous reflection on implementation progress and learning.

To bridge the gap between the impact questions CTR is asking and what the implementers are observing, learning, and reporting, CTR should consider the following steps:

1. Examine the array of surveys and their components and determine or prioritize which elements are most relevant, and how they match up with CTR's monitoring and learning priorities.
2. Establish clear indicators and data-collection processes for implementers to integrate into activities that match CTR's priorities and impact-outcome needs. This will inform the reporting requirements from implementers as they collect data and report on progress.

3. Conduct a stakeholder mapping and analysis to understand the M&E needs of different stakeholders (i.e., facilities, associations, implementer, program officers, and others) and how to tailor metrics and surveys to collect relevant data.
4. Assess how data, analysis, and reports are stored and shared to better inform programming and improve impact over time.